

*Ana Sesartic Petrus und Matthias Töwe*

# Forschungsdatenmanagement an der ETH Zürich: Ansätze und Wirkung

**Zusammenfassung:** Nicht nur die komplexer werdenden Arbeitsprozesse in der Forschung, sondern auch Vorgaben von Geldgebern machen ein professionelles Datenmanagement erforderlich. Die ETH Zürich unterstützt ihre Forschenden auf vielfältige Weise. Der Artikel beschreibt konkrete technische Lösungen die im Einsatz sind, sowie das Beratungs- und Kursangebot der ETH Zürich. Statistiken zeigen, welche Kundengruppen erreicht werden und wo noch Handlungsbedarf besteht.

**Schlüsselwörter:** Forschungsdatenmanagement, Datenlebenszyklus, Dienstleistungen

## Research Data Management at ETH Zurich: Approaches and Their Effects

**Abstract:** Not only the increasingly complex methods in research, but also the policies implemented by funding agencies make professional data management a necessity. ETH Zurich supports its researchers in many ways. The article describes concrete technical solutions in use, as well as the consulting and course offerings of ETH Zurich. Survey data show which customer groups the available services already reach and where further action is still required.

**Keywords:** Research data management, Data life-cycle, Services

## 1 Einleitung

Die Notwendigkeit, Forschungsdaten zu managen ist nicht erst durch die umfassende Digitalisierung des Forschungsprozesses entstanden. Die Digitalisierung erhöht aber den Druck sowohl aus quantitativer als auch aus qualitativer Sicht. In bestimmten Fächern werden Forschungsdaten mit einer Geschwindigkeit erzeugt, die allein schon die ausreichend schnelle Speicherung zur Herausforderung macht (Bornmann and Mutz 2015; Van Noorden 2014). Rein manuelle Methoden kommen dementsprechend für die Dokumentation und Organisation der Daten nicht mehr in Frage.

Neben diesem Druck schafft die Digitalisierung auch erstmals die Möglichkeit, Forschungsdaten in großem Stil in nutzbarer Form auszutauschen, statt sie nur z.B. in Tabellenform im Anhang eines gedruckten Aufsatzes zu veröffentlichen. Entsprechend hoch sind je nach Fach die Erwartungen an die Bereitstellung von Forschungsdaten für die Community.

Es ist allerdings wichtig, sich bewusst zu machen, dass Forschungsdaten keineswegs immer einen großen Umfang haben müssen oder ihrem unstrukturierten Charakter nach als „Big Data“ zu betrachten sind. Selbst an einer technisch-naturwissenschaftlich ausgerichteten Hochschule wie der ETH Zürich trifft dies nur auf bestimmte Forschungsgruppen zu, wobei ein Schwerpunkt in denjenigen Departementen liegt, die sich mit datenintensiven Gebieten der Lebenswissenschaften oder mit Klimabeobachtung und –modellierung befassen. Daneben gibt es den bekannten „Long Tail“ (Wallis, Rolando, and Borgman 2013) von Forschungsdaten mit einer breiten Verteilung vieler kleinerer Datensammlungen über die verschiedensten Fachgebiete hinweg. Das Forschungsdatenmanagement sollte nicht primär als technisches Problem verstanden werden: Technische Lösungen sind in manchen Bereichen unverzichtbar und können in vielen Fällen für Verbesserung sorgen, doch ist der angemessene Umgang mit Forschungsdaten vor allem eine organisatorische und Management-Aufgabe.

Es bestehen große Unterschiede in der Arbeitsweise der jeweiligen Forschungsgruppen sowie hinsichtlich der Zusammenarbeit innerhalb der Fachcommunity: Selbstverständliche Gewohnheiten aus einem Fach haben u.U. überhaupt keine Tradition in anderen Disziplinen. Dies gilt ganz besonders für die Frage nach dem offenen Austausch von Forschungsdaten. Je nach Fach stellen Zeitschriftenherausgeber in unterschiedlichem, aber stetig zunehmenden Maß Anforderungen an die Bereitstellung von Daten, die Veröffentlichungen zugrunde liegen (z.B. (Springer Nature 2018)). Die Regelungen seitens der Geldgeber für die Forschung haben dagegen fachunabhängig Gültigkeit. Entsprechende Vorgaben machen z.B. der Schweizerische Nationalfonds SNF (Schweizerischer Nationalfonds 2017) oder das EU Forschungsrahmenprogramm Horizon 2020 (European Commission 2018), die konkret auch die Einreichung von Datenmanagementplänen verlangen. Ihr Hauptinteresse ist es, die Nutzbarkeit von mit öffentlichen Geldern gesammelten Daten sicherzustellen, um idealerweise das Wissenschaftssystem insgesamt zu beflügeln. Die Interessen der Geldgeber stimmen dabei vor allem kurzfristig nicht automatisch mit jenen der einzelnen Forschenden überein. Diese Heterogenität der konkreten Arbeitsweisen bildet eine der grundlegenden Herausforderungen, mit denen sich Hochschulen konfrontiert sehen: Infrastrukturen, Prozesse und Richtlinien müssen für die ganze Hochschule und ihre Fächer zur Verfügung stehen und anwendbar sein. Dies stellt hohe Anforderungen an diejenigen Personen und Stellen, die innerhalb der Hochschule Forschende unterstützen.

Sofern anerkannte fachspezifische Plattformen und Repositorien bestehen, brauchen die einzelnen Institutionen unter Umständen nur subsidiäre Angebote aufzubauen für Anwendungsfälle oder Fächer, die nicht schon anderweitig angemessen abgedeckt sind.

Im Folgenden wird dargestellt, wie an der ETH Zürich die Unterstützung für das Datenmanagement auf verschiedenen Ebenen aufgebaut wurde. Zudem wird gezeigt, welche Forschenden damit bisher erreicht werden konnten. Die Autoren schreiben in ihrer Rolle als Mitarbeitende der Gruppe Forschungsdatenmanagement und Datenerhalt der ETH-Bibliothek und beziehen Angebote von Partnern außerhalb der Bibliothek ein. Trotz mehrjähriger Erfahrung ist dieser Überblick/Bericht als Momentaufnahme zu verstehen, da kontinuierlich Anpassungen und Ergänzungen vorgenommen werden.

## 2 Situation an der ETH Zürich

Die ETH Zürich ist mit 20'600 Studierenden und 530 Professorinnen und Professoren (Stand 2017<sup>1</sup>) die größte Einrichtung im ETH-Bereich<sup>2</sup> der von der Schweizerischen Eidgenossenschaft getragen wird. Zum ETH-Bereich gehören daneben die zweite vom Bund finanzierte Technische Hochschule EPFL in Lausanne sowie die vier Forschungsanstalten Eawag, Empa, PSI und WSL mit jeweils unterschiedlicher Ausrichtung.<sup>3</sup>

Innerhalb der ETH Zürich wird die Unterstützung für das Forschungsdatenmanagement vor allem vom Bereich des Vizepräsidenten für Personal und Ressourcen getragen, bei dem die ETH-Bibliothek und die Informatikdienste der ETH Zürich als Abteilungen angesiedelt sind.

Es ist Aufgabe der ETH-Bibliothek, die Forschenden an der ETH Zürich vom Anfang des Datenlebenszyklus an zu unterstützen. Dies beginnt mit allgemeiner Beratung bezüglich Datenmanagementplänen, der Compliance mit den aktuellen Vorgaben der Geldgeber bis zur Publikation der Daten, der DOI-Vergabe und Langzeitarchivierung.

Was die Unterstützung im aktiven Forschungsdatenmanagement (FDM) angeht, d.h. beim Datenmanagement während der aktiven Erfassung, Erstellung und Analyse der Daten, so wird zusätzliche Expertise von außerhalb der ETH-Bibliothek benötigt. Diese Aufgaben werden von der Sektion Scientific IT Services (SIS) der zentralen Informatikdienste der ETH Zürich erbracht (ETH Zürich 2018). Diese unterstützt seit Jahren Forschungsgruppen und hat seit Anfang 2018 den offiziellen Auftrag der Hochschulleitung, Dienstleistungen und Lösungen bezüglich des aktiven Datenmanagements an der ETH Zürich insgesamt zu entwickeln/anzubieten. Dies drückt sich v.a. in der Bereitstellung des ETH Research Data Hub (ETH-RDH)<sup>4</sup> aus. Dabei handelt es sich um die zentral gehostete Version des in-house entwickelten Laborinformationsmanagementsystems und

---

<sup>1</sup> <https://www.ethz.ch/de/die-eth-zuerich/portraet/die-eth-zuerich-in-zahlen.html>.

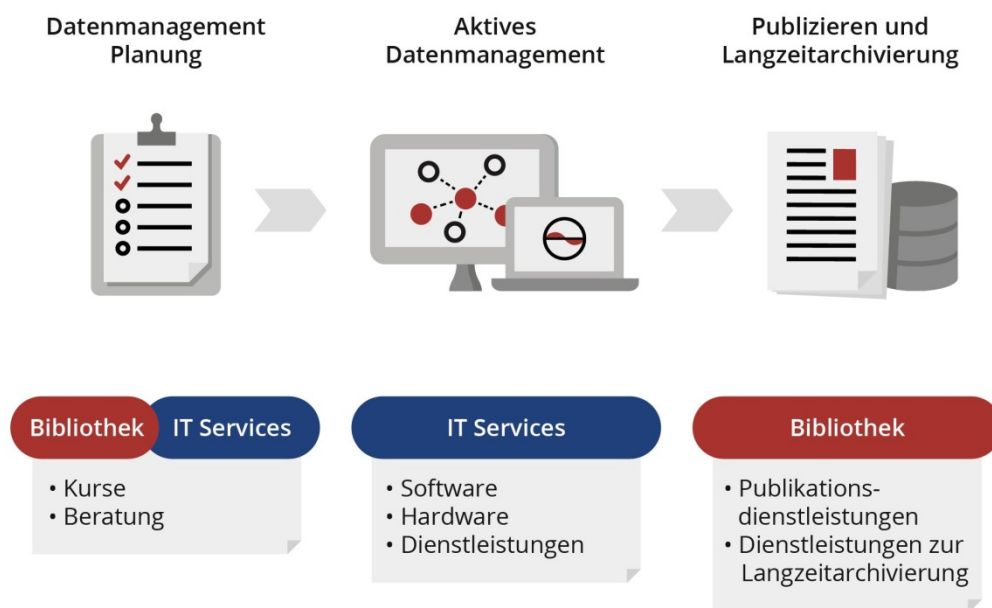
<sup>2</sup> <https://www.ethrat.ch/de/eth-bereich/auf-einen-blick>.

<sup>3</sup> <https://www.ethrat.ch/de/node/235>.

<sup>4</sup> <https://labnotebook.ch>.

Elektronischen Laborbuchs openBIS der ETH-Zürich. Dieses kann als Instanz pro Forschungsgruppe installiert und flexibel an die Bedürfnisse der jeweiligen Gruppe angepasst werden. Parallel werden die Forschenden in Kursen und persönlichen Beratungen durch SIS in ihrer Arbeit unterstützt (für Details siehe folgendes Kapitel zu ETH Research Data Hub).

Um eine bessere Koordination zwischen der ETH-Bibliothek und den Scientific IT Services zu gewährleisten (siehe Abb. 1 für einen Kurzüberblick) und eine zentrale Anlaufstelle für Forschende zu schaffen, wurde im Sommer 2017 die Website [www.ethz.ch/researchdata](http://www.ethz.ch/researchdata) ins Leben gerufen, die gemeinsam von der ETH-Bibliothek und SIS bewirtschaftet wird. Dort finden Forschende weiterführende Informationen zu jeder Station des Datenlebenszyklus sowie zu den zuständigen Kontaktstellen. Außerdem finden sich dort die Neuigkeiten zum Thema, z.B. demnächst stattfindende Kurse und Workshops, oder wichtige Neuigkeiten der Geldgeber. Falls ein Kunde unsicher ist, an wen er sich wenden soll, bietet eine Anfrage an die zentrale Mailadresse Abhilfe, von welcher aus der Kunde an die richtige Fachstelle verwiesen wird.



**Abb. 1:** Übersicht über die von der ETH-Bibliothek und den Scientific IT Services der ETH Zürich angebotenen Dienstleistungen entlang des Datenlebenszyklus

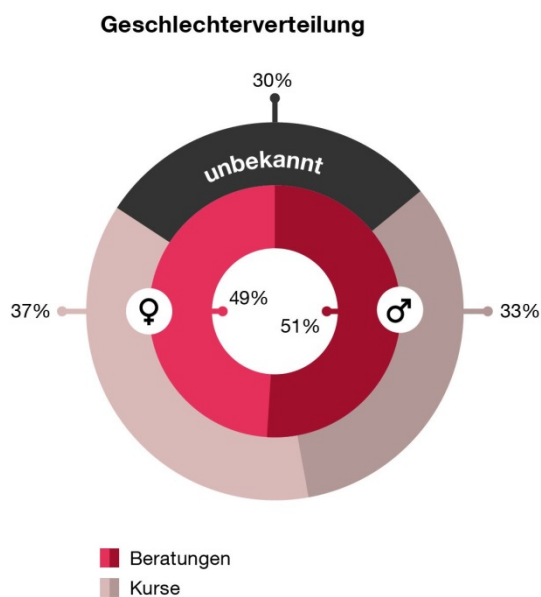
Die derzeitige Aufgabenteilung ist stark von der Ausgangslage innerhalb der Hochschule geprägt und erhebt selbstverständlich keinen Anspruch auf Übertragbarkeit. Angesichts der Dynamik der Aufgaben im Forschungsumfeld ist weitgehend offen, wie lange eine einmal etablierte Lösung

tatsächlich Bestand haben wird bzw. wie schnell die Beteiligten auf neue Bedürfnisse reagieren müssen.

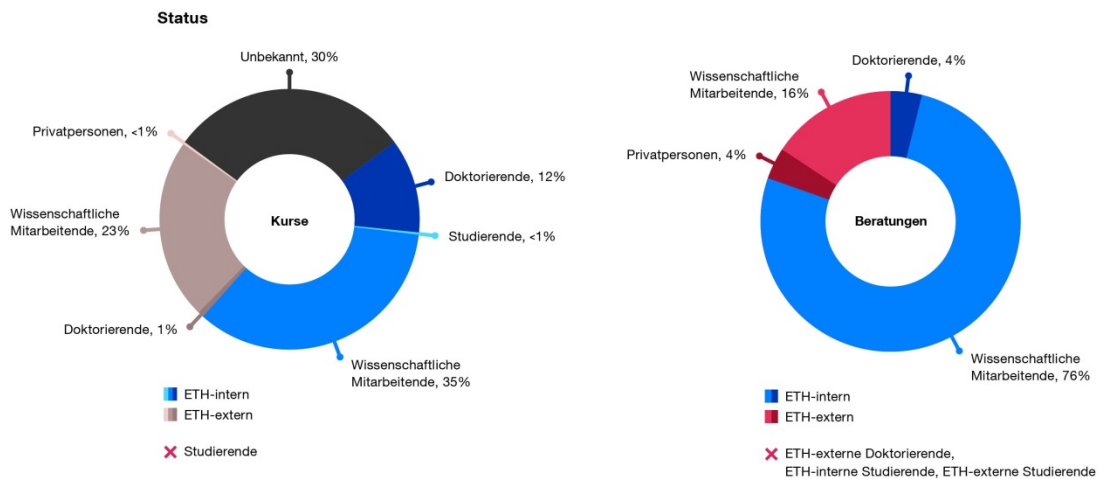
## 2.1 Kurse und Beratungen

Aktuell bietet die ETH-Bibliothek zum Thema FDM diverse Kursformen (Informationsveranstaltungen, Workshops) sowie personalisierte Kurse für einzelne Forschungsgruppen an. Ziel dieser Kurse ist es, das Bewusstsein für bestehende Anforderungen und die Vorteile des FDM zu schärfen, Konzepte für die Umsetzung in der Praxis zu vermitteln, die an der ETH Zürich angebotenen Dienste und Tools bekanntzumachen, und den Austausch über die Forschungsgruppen hinweg zu fördern. Die Forschenden sollen in die Lage versetzt werden, bewusste Entscheidungen über ihre Daten zu treffen, da sie selbst über das größte Expertenwissen verfügen.

Die Kurse werden kostenlos angeboten und sind für jedermann zugänglich, wobei die angestrebte Zielgruppe die Angehörigen der ETH Zürich sind. Zusätzlich haben sich die Kurse und Beratungen als hervorragende Marketinginstrumente erwiesen, da fast nach jedem Kurs Anfragen von Teilnehmenden für eine maßgeschneiderte Veranstaltung für ihre Forschungsgruppe oder ihr Institut eingehen.

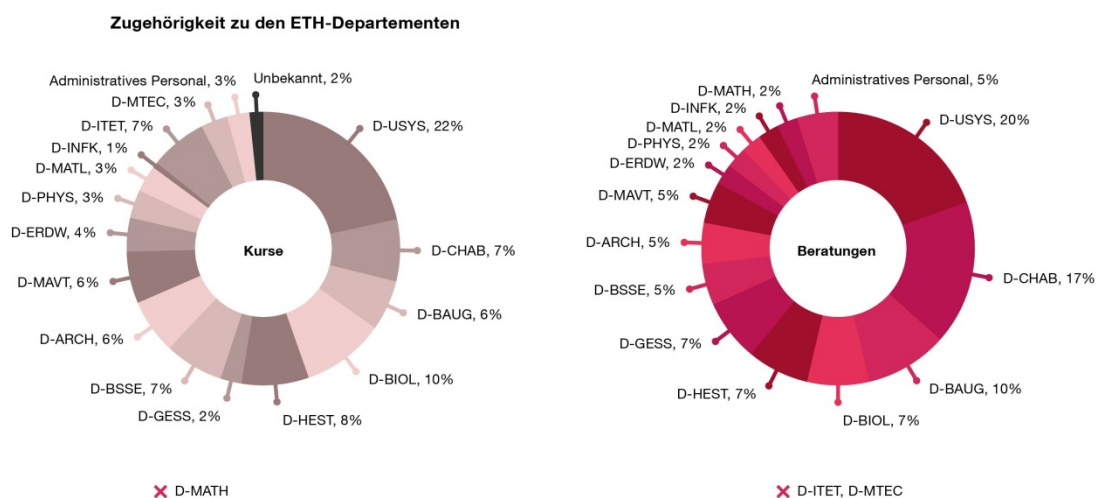


**Abb. 2:** Geschlechterverteilung unter den Kurs- und Beratungsteilnehmern



**Abb. 3:** Anteil von ETH-Angehörigen und externen Kurs- und Beratungsteilnehmern, sowie ihre Karrierestufe

Die folgenden Zahlenangaben sowie die Werte in den Abbildungen sind gerundet, so dass die Summe nicht in jedem Fall 100% beträgt. Obwohl an der ETH Zürich der Frauenanteil mit 35% immer noch relativ tief liegt (ETH Zürich 2017) werden unsere Kurse gleichmäßig von den beiden Geschlechtern besucht, mit leicht höherem Frauenanteil (53%, siehe Abb. 2). Obwohl die Kurse hauptsächlich auf ETH-Forschende ausgerichtet sind, kommen 34% der Teilnehmenden unserer Veranstaltungen von anderen Hochschulen und Institutionen (siehe Abb. 3), wobei diese hauptsächlich wissenschaftliche Mitarbeiter sind. Wie man aus der Abb. 3 außerdem sieht, erreichen wir mit unseren Kursen meist erfahrene Forschende. Für uns überraschend war, dass nur 18% der Kursbesucher Doktorierende waren. Da bei gewissen Veranstaltungen nur die Anzahl Teilnehmer erfasst wurde und nicht ihr Geschlecht, sind diese in den Abb. 2 und 3 als „unbekannt“ gekennzeichnet.



**Abb. 4:** Verteilung der Kunden nach Zugehörigkeit zu den ETH-Departementen

Schaut man sich die Verteilung auf die Departemente der ETH an (Abb. 4), fällt auf, dass 23% dem Departement D-USYS (d.h. den Umweltsystemwissenschaften) angehören. Diese Teilnehmergruppe ist sensibilisiert und offen für die Problematik des FDM. Da die Communities in dieser Disziplin häufig mit einzigartigen (z.B. Naturbeobachtungen) oder großen Datenmengen (z.B. Outputs von Klimamodellen) arbeiten, zeigt sich bei ihnen schon lange ein ausgeprägtes Verständnis für die Notwendigkeit eines funktionierenden Datenmanagements. Die Departemente Biologie, Biosysteme und Gesundheitswissenschaften kommen gemeinsam auf 25%, vermutlich, da sie ebenfalls schon länger durch die Art und Menge ihrer Daten mit drängenden Anforderungen an das Datenmanagement konfrontiert sind.

Interessanterweise hält sich die Teilnahme der Physiker (4%), Materialwissenschaftler (3%) und Informatiker (1%) dagegen in engen Grenzen. Gründe dafür können vielfältiger Natur sein. Gemäß Aussagen der wenigen Kursteilnehmenden kann nur vermutet werden, dass in diesen Departementen die Ansicht stärker verbreitet ist, bereits über genügend Wissen in diesem Bereich zu verfügen. Von Einzelnen wurde zudem geäußert, dass v.a. seitens der Gruppenleitung dem Forschungsdatenmanagement keine große Bedeutung beigemessen werde. Es lässt sich nicht belegen, ob dies Ausnahmen sind und ob sie mit der fachlichen Ausrichtung in Verbindung stehen. Eine interessante Beobachtung ist auch, dass wir bisher keinen einzigen Teilnehmer vom Departement Mathematik hatten. Hier ist zu bedenken, dass speziell bei theoretisch arbeitenden Mathematikern tatsächlich über ihre Publikationen hinaus wenig oder keine greifbaren Daten anfallen.



**Abb. 5:** Anzahl Teilnehmender an Beratungen und Kursen

Insgesamt wurden im Zeitraum vom September 2015 bis August 2018 über 1300 Teilnehmende in verschiedenen Veranstaltungsformaten über das Thema FDM informiert und geschult (siehe Abb. 5). Da diese einen sehr heterogenen Hintergrund aufweisen, variieren auch ihr Wissensstand und ihre Bedürfnisse relativ stark. Alle waren sich jedoch der Problematik bewusst und suchten nach Lösungen.

Bei Beratungen zu den Themen DMP und FDM allgemein sind ab Juni 2017 Daten verfügbar. In dieser Zeit nahmen bis August 2018 46 Personen insgesamt 92 Beratungen in Anspruch (viele haben sich mehrmals gemeldet). Obwohl das Beratungsangebot hauptsächlich auf ETH-interne Personen zugeschnitten ist, wird es auch von Angehörigen des gesamten ETH-Bereichs genutzt. Ihr Anteil beträgt 20%. Wie schon bei den Kursen, zeigt sich auch bei den Beratungen ein ausgeglichenes Geschlechterverhältnis (48% Frauen, 52% Männer, siehe Abb. 2), wobei aber der Anteil der ETH-Forschenden mit 80% deutlich höher liegt als bei den Kursen. Interessanterweise kam nur gerade eine Anfrage von einem Doktoranden. Alle anderen Kunden waren entweder bereits Postdoktoranden, wissenschaftliche Mitarbeiter oder Professoren.

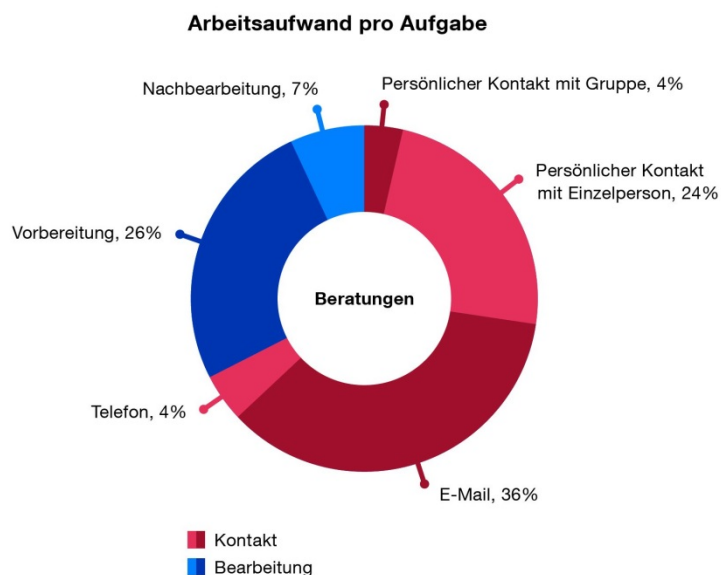
Die Verteilung nach Departementen ist insofern ähnlich, als dass 26% der Anfragen vom D-USYS kamen. Aber im Gegensatz zu den Kursen liegt das Departement Chemie und Angewandte Biowissenschaften (D-CHAB) fast gleichauf. Die Biologie, Biosysteme und Gesundheitswissenschaften kommen gemeinsam auf 26%, ähnlich wie bei den Kursen. Insgesamt wurden nur Forschende aus 10 der insgesamt 16 ETH-Departemente (siehe Tabelle 1) beraten.



| Kürzel | Departement                                |
|--------|--|
| D-ARCH | Architektur                                |
| D-BAUG | Bau, Umwelt und Geomatik                   |
| D-BSSE | Biosysteme                                 |
| D-INFK | Informatik                                 |
| D-ITET | Informationstechnologie und Elektrotechnik |
| D-MAVT | Maschinenbau und Verfahrenstechnik         |
| D-MATL | Materialwissenschaft                       |
| D-BIOL | Biologie                                   |
| D-CHAB | Chemie und Angewandte Biowissenschaften    |
| D-MATH | Mathematik                                 |
| D-PHYS | Physik                                     |
| D-ERDW | Erdwissenschaften                          |
| D-HEST | Gesundheitswissenschaften und Technologie  |
| D-USYS | Umweltsystemwissenschaften                 |
| D-MTEC | Management, Technologie und Ökonomie       |
| D-GESS | Geistes-, Sozial- und Staatswissenschaften |

**Tab. 1:** Auflistung der Departemente an der ETH Zürich<sup>5</sup>

Die meisten Beratungsanfragen erreichen die Gruppe Forschungsdatenmanagement und Datenerhalt via E-Mail und stammen von Einzelpersonen (Siehe Abb. 6). Gewisse Anfragen erfordern zusätzliche Recherchen und Vorbereitung um sie zu beantworten, die meisten können aber rasch gelöst werden. Grundsätzlich wird darauf geachtet, die Kundenanfragen innerhalb von 24 Stunden zu beantworten. Wenn ersichtlich wird, dass dies nicht möglich ist, werden die Kunden darauf aufmerksam gemacht.



**Abb. 6:** Arbeitsaufwand pro Beratungsaufgabe

<sup>5</sup> <https://www.ethz.ch/de/die-eth-zuerich/organisation/departemente-und-kompetenzzentren/departemente.html>.

Beratungsanfragen drehen sich hauptsächlich um die Thematik der Datenmanagementpläne. Dies beginnt mit Fragen dazu, wie diese am besten auszufüllen sind, um den Anforderungen der Forschungsförderer (in erster Linie des Schweizerischen Nationalfonds SNF) zu entsprechen und reicht bis zur Bitte um ein Gegenlesen des Entwurfs des DMP vor dem Einreichen des Gesuchs. Die Forschenden sind durch die Vorgaben häufig verunsichert, weil die notwendigerweise allgemeinen Vorgaben auf den ersten Blick nicht zu ihrer eigenen Arbeitsweise zu passen scheinen. Sobald vermittelt werden kann, dass Forschende selbst die Anforderungen interpretieren dürfen und sollen, wird die Aufgabe für sie wesentlich klarer. Liest man schließlich ihre DMPs, so sieht man, dass diese meist plausibel sind und zumindest die Anforderungen an die erste einzureichende Fassung ohne Weiteres erfüllen.

Besonders viele Fragen warf die ursprüngliche Ankündigung des SNF auf, die die Forschenden zur Einreichung von DMPs verpflichtete und die DMPs sehr eng mit dem Thema „Open Research Data“ verknüpfte. Selbstverständlich besteht hier ein sachlicher Zusammenhang, doch hat die Betonung der Forderung nach Offenlegung von Daten unnötigen Widerstand und Verwirrung provoziert. Bei genauerer Betrachtung der Vorgaben besteht die Mindestanforderung darin, lediglich diejenigen Daten offenzulegen, die einer Publikationen zugrunde liegen. Damit ist natürlich die Diskussion eröffnet, wo hier die Grenze zu ziehen ist: Welche Daten gehören zu einer Publikation und welche können z.B. als reine Zwischenprodukte gelöscht und bei Bedarf wieder erzeugt werden? Wie auch andere Fragestellungen zum Umgang mit Forschungsdaten kann auch diese Frage letztlich nur von den Mitgliedern der jeweiligen Fachcommunities beantwortet werden, die für sich klären müssen, welches ihre Erwartungen aneinander sind. Es ist sicher ein Verdienst der Vorgaben des SNF, diese Diskussion in der Schweiz zumindest intensiviert zu haben.

Des Weiteren beziehen sich viele Fragen von Forschenden auf die Thematik der Langzeitarchivierung und Ablage der Forschungsdaten. Dies gilt nicht nur für die Ablage der aktuellen Forschungsausgaben in unserem Hochschulrepositorium, der Research Collection (Details siehe weiter unten), sondern auch für die Archivierung von Unterlagen emeritierter Professoren. Deshalb arbeiten wir auch eng mit dem Hochschularchiv (HSA) der ETH Zürich zusammen, das Teil der ETH-Bibliothek ist. Gemäß dem Reglement für das Archiv der ETH Zürich, Art. 3, (ETH Zürich 2015), gestützt auf das Schweizerische Bundesgesetz über die Archivierung, Art. 6 (Schweizerische Eidgenossenschaft 1999), werden Professoren vor der Emeritierung dazu aufgerufen, ihre analogen und digitalen Unterlagen dem HSA anzubieten. Der Ordnungszustand der abgelieferten Unterlagen schwankt jedoch stark und es besteht häufig eine Vermischung von Geschäftsunterlagen und Forschungsdaten. Bei gewissen Unterlagen fehlt der Kontext, Dateiformate sind z.T. veraltet und unlesbar, und Metadaten fehlen sowohl bei digitalen als auch analogen Dokumenten. Spätestens hier zeigt sich, dass Professuren in

aller Regel nicht über Knowhow zum Thema Records Management von Geschäftsunterlagen verfügen, während der Informationsstand und das Engagement beim Thema Forschungsdatenmanagement stark schwanken.

Um diesen Zustand zu verbessern, muss man bereits bei jungen Forschenden mit der Ausbildung und Sensibilisierung ansetzen. Wenn Forschungsdatenmanagement nicht als Führungsaufgabe wahrgenommen und somit nicht „zur Chefsache“ erklärt wird, (die es gemäss den Richtlinien für Integrität in der Forschung der ETH Zürich (ETH Zürich 2011) eigentlich bereits sein sollte), werden viele emeritierte Professoren weiterhin ihre Unterlagen und damit verbundene Daten unstrukturiert im Hochschularchiv ablagern, in der Hoffnung, dass die Unordnung von Archivaren schon aufgeräumt werden wird. Letzteres ist jedoch nur beschränkt möglich, da diese weder über das dafür nötige disziplinspezifische Fachwissen noch über ausreichende Personalressourcen verfügen.

Forschungsdatenmanagement an sich ist keine neue Erfindung und war bei gewissen Professuren schon länger „Chefsache“ – mit mehr oder weniger Erfolg. Viele Forschende haben festgestellt, dass effiziente Forschung nur dann möglich ist, wenn Daten und Metadaten gepflegt werden. Wie überzeugende Beispiele zeigen, hat FDM an der ETH Zürich schon stattgefunden, lange bevor Geldgeber das Thema prominent adressiert haben und Datenmanagementpläne verlangten.

Ein Beispiel (Sesartić, Fischlin, and Töwe 2016) stammt von der Gruppe für terrestrische Ökosysteme an der ETH Zürich unter der Leitung von Professor Andreas Fischlin. Die interdisziplinär forschende Gruppe nutzte verschiedene Datenquellen von hoher Komplexität. Deshalb entwickelte die Gruppe unter der Leitung von Prof. Fischlin, seit ihrer Gründung im Jahre 1988, ein Konzept zum Management ihrer Forschungsdaten. Dazu gehörten u.a. Vorgaben über Metadaten, Dateinamen, Dateiablage, die Benutzung der gemeinsamen Literaturliteraturdatenbank, sowie den Ablauf für die Archivierung von Daten – insbesondere wenn ein Projekt abgeschlossen wurde. Es wurden konsequent die modernsten verfügbaren Methoden der jeweiligen Zeit genutzt und die archivierten Daten auf die jeweils aktuellen Datenträger kopiert – z.B. von Magnetbändern auf CDs usw.

Das Konzept wurde über Jahrzehnte erfolgreich umgesetzt, zeigte aber auch Grenzen in der Praxis auf. Trotz sorgfältiger Planung gemäß damaligen „best practices“, machten die Materialalterung und die sogenannte „Software-Erosion“ (das Veralten der benötigten Software) den Forschenden einen Strich durch die Rechnung. Schneller als erwartet wurden z.B. gewisse Lesegeräte durch Alterung unbrauchbar oder beschädigten sogar die Datenträger, oder die Daten konnten aufgrund der genannten Software-Erosion nur noch in Rohform gerettet werden (Sesartic 2017). Es wurde versucht, eigene Software auf neuer Hardware lauffähig zu halten, doch bei kommerzieller Software war dies nicht möglich. Da zudem bekannt war, dass die Software nicht hardwareunabhängig

ausgeführt werden konnte, wurden routinemäßig Referenzoutputs von Modellierungen archiviert, um die Vergleichbarkeit zu gewährleisten.

### **3 Konkrete technische Lösungen**

#### **3.1 Openbis: Von der Applikation zum Service**

Während der laufenden wissenschaftlichen Arbeit kommt dem Management von Forschungsdaten entscheidende Bedeutung zu. In dieser Phase können und müssen wichtige Kontextinformationen erfasst werden, die später nicht, nur unvollständig oder mit großem Aufwand ergänzt werden können. Gleichzeitig besteht je nach Anzahl der anfallenden Objekte der Bedarf, sie in vordefinierten Prozessen automatisch zu strukturieren und zu annotieren.

An der ETH Zürich wurden die Anforderungen an Systeme für das aktive Datenmanagement zunächst vor allem von Forschenden der Lebenswissenschaften formuliert. Ab 2007 wurde die Open Source Anwendung openBIS entwickelt, die heute von den Scientific IT Services der ETH Zürich unterhalten wird (Barillari et al. 2016). Sie ist als Plattform konzipiert, um unterschiedlichste Werkzeuge zum Verarbeiten von Daten einbinden zu können. Die Funktionalitäten von openBIS konzentrieren sich somit auf das Management der beteiligten Komponenten und Daten, auf die sie angewendet werden. Für die eigentliche wissenschaftliche Datenverarbeitung stützt sich openBIS dagegen auf existierende, anerkannte Werkzeuge. Da bei den Lebenswissenschaften der Bedarf am größten war, wurden zunächst Tools integriert, die in diesen Fächern verbreitet sind. Bis heute hat sich openBIS zu einem umfassenden Laborinformationsmanagementsystem entwickelt, das auch die Funktion eines Elektronischen Laborjournals erfüllt.<sup>6</sup> Es ermöglicht die nachvollziehbare Durchführung einer Vielzahl von Aktivitäten im Forschungsprozess wie beispielsweise die Verwaltung physischer Proben, die Annotation von Daten in verschiedenen Stadien der Bearbeitung, die Einrichtung automatisierter Bearbeitungsworkflows oder die Verlinkung von extern vorgehaltenen großen Datensätzen. Die große Stärke dabei ist die dokumentierte Vernetzung aller Objekte, deren Abhängigkeiten auf anderen Wegen nur noch mit großem Aufwand zu handhaben sind.

In den vergangenen Jahren wurde openBIS einzeln pro Forschungsgruppe installiert und gemäß deren Anforderungen konfiguriert. Neu haben die Informatikdienste der ETH Zürich den Auftrag, eine Basisinfrastruktur für das aktive Datenmanagement für die ganze Hochschule anzubieten. Gegenwärtig laufen die Vorbereitungen, um auf der Basis von openBIS den zentral gehosteten Service ETH Research Data Hub für das Datenmanagement in laufenden Forschungsprojekten bereitzustellen. Dieser Service ist für Forschungsgruppen gedacht, die nur eine begrenzte Zahl von

---

<sup>6</sup> <https://labnotebook.ch>.

Objekten verwalten müssen und keine speziell auf sie zugeschnittene Konfiguration benötigen. Für andere Gruppen wird weiterhin die Möglichkeit bestehen, openBIS als individuelle Installation einrichten zu lassen. Grundsätzlich kann die Anwendung auch ohne weitere Unterstützung als Open Source Software bezogen und genutzt werden.

Es sind Überlegungen im Gange, openBIS an die Research Collection anzubinden, um ausgewählte Daten aus der Datenmanagementplattform heraus zu publizieren und zitierbar zu archivieren.

### **3.2 Research Collection und ETH Data Archive**

Die ETH-Bibliothek unterstützt sowohl die Veröffentlichung von Forschungoutput der ETH Zürich als auch die Langzeitarchivierung von veröffentlichtem und unveröffentlichtem Material. Bis Anfang 2017 betrieb die ETH-Bibliothek die E-Collection als institutionelles Repositorium für die verpflichtende Ablieferung elektronischer Dissertationen sowie für die Zweitveröffentlichung von Publikationen und die Bereitstellung von grauer Literatur. Damals war die Veröffentlichung von Forschungsdaten auf diesem Weg noch nicht vorgesehen. Auf der gleichen technischen Basis wie die E-Collection (Fedora<sup>7</sup>) wurde parallel die Hochschulbibliographie E-Citations unterhalten. Beide Gefäße wurden trotz funktionaler Querverbindungen separat betrieben. So war es möglich, eine Volltextablieferung an die E-Collection gleichzeitig auch an E-Citations zu melden und umgekehrt einem bibliographischen Eintrag in E-Citations einen Volltext zum Laden in die E-Collection anzuhängen. Im Hintergrund blieben die Systeme jedoch getrennt.

Mitte 2017 wurden E-Citations und E-Collections abgelöst. Seither bietet die ETH-Bibliothek die Research Collection<sup>8</sup> auf der Basis von DSpace<sup>9</sup> als Repositorium mit dreifacher Funktion an: als institutionelles Repositorium der ETH Zürich für Erst- und Zweitveröffentlichungen von Dokumenten, als Hochschulbibliographie und als Forschungsdatenrepositorium (Hirschmann 2018). Letztere Funktion dient zur Ablage und zitierbaren Veröffentlichung von Forschungsdaten, sei es als Zusatzmaterial zu Publikationen oder als eigenständige Objekte. Dadurch kommt der Research Collection eine umfassende Bedeutung als Publikationsplattform für den Forschungoutput der ETH Zürich zu. Kunden können alle Funktionen unter der gleichen Oberfläche und mit einheitlicher Bedienung nutzen.

Neben der unmittelbaren Veröffentlichung besteht auch die Möglichkeit, den Zugriff erst nach einer Sperrfrist freizugeben oder dauerhaft zu beschränken. Dank dieser Option kann die Research Collection bis zu einem gewissen Grad auch eine Funktion in frühen Phasen des Lebenszyklus der

---

<sup>7</sup> <https://duraspace.org/fedora>.

<sup>8</sup> <https://www.research-collection.ethz.ch>.

<sup>9</sup> <https://duraspace.org/dspace>.

Forschungsdaten übernehmen, wenn andere Möglichkeiten zur strukturierten Ablage von Daten fehlen. Der Schwerpunkt der Research Collection liegt aber klar auf der öffentlichen Bereitstellung, denn die Metadaten der hochgeladenen Objekte werden in jedem Fall veröffentlicht.

Für ETH-Angehörige besteht keine Verpflichtung, die Research Collection für Forschungsdaten zu nutzen. Es steht ihnen frei, nationalen oder internationalen Repositorien den Vorzug zu geben, die in ihrer jeweiligen Fachcommunity anerkannt sind und eine gewisse Nachhaltigkeit erwarten lassen.

Der Inhalt der Research Collection wird automatisch zum Zweck der digitalen Langzeitarchivierung an das ETH Data Archive übergeben. Dabei handelt es sich um die zentrale Lösung der ETH Zürich für die Langzeitarchivierung auf der Basis der Anwendung Rosetta der Firma Ex Libris.<sup>10</sup> Einige Jahre lang wurden über das ETH Data Archive auch Workflows zur manuellen Ablage und Veröffentlichung von Forschungsdaten abgewickelt. Kunden luden dazu Datenpakete über einen Web-Client ins System und die Metadaten wurden anschließend in das öffentliche Discovery Tool „Wissensportal“ (heute „Suchportal“<sup>11</sup>) der ETH-Bibliothek exportiert (Applikation Primo, Fa. Ex Libris<sup>12</sup>).

Diese Phase war nützlich, um Erfahrungen mit den Bedürfnissen von Kunden zu sammeln, die über den Kreis der in die Entwicklung einbezogenen Forschenden hinaus gingen. Es zeigten sich dabei klar Limitierungen in der Flexibilität, die der Funktion des ETH Data Archive als digitales Langzeitarchiv geschuldet sind: Ein Langzeitarchiv soll ja im Normalfall nach dem Ingest gerade keine „einfachen“ Manipulationen durch Endnutzer mehr zulassen. Mit dem Start der Research Collection wurde das ETH Data Archive bis auf wenige Spezialfälle weitgehend von den Aufgaben der Datenveröffentlichung entlastet.

Ein großer Teil der Anforderungen dieser späteren Phasen des Datenlebenszyklus' kann mit den an der ETH Zürich vorhandenen Lösungen angemessen abgedeckt werden. Noch offen ist eine Lösung zur Registrierung von sehr großen Datensätzen, die nicht über den Webbrowser geladen werden können und allenfalls nicht einmal online vorgehalten werden können. Zwar ist eine Aufnahme der Metadaten bereits heute möglich, jedoch ohne einen standardisierten Prozess für die Datenablage und –nutzung. Ein weiteres Ziel ist die Anbindung von Research Collection bzw. ETH Data Archive an bestehende Plattformen für das aktive Forschungsdatenmanagement wie openBIS, um einen bruchlosen Transfer von Daten aus der Arbeitsumgebung der Forschenden an die Publikationsplattform bzw. das Langzeitarchiv zu ermöglichen.

---

<sup>10</sup> <https://www.exlibrisgroup.com/products/rosetta-digital-asset-management-and-preservation>.

<sup>11</sup> <https://search.library.ethz.ch>.

<sup>12</sup> <https://www.exlibrisgroup.com/products/primo-library-discovery>.

### 3.3 Docuteam packer und feeder

Bereits in einer frühen Phase des Projekts zum Aufbau des ETH Data Archive war klar, dass Forschende Bedarf nach einer einfach zu bedienenden Lösung hatten, um Daten frühzeitig lokal zu organisieren und mit Metadaten zu versehen. Diese sollte nicht eine umfassende Anwendung für das Datenmanagement wie z.B. ein Laborinformationsmanagementsystem (LIMS) ersetzen, sondern es den Forschenden ermöglichen, Angaben zum Entstehungskontext der Daten erfassen und mit den Daten zusammen für die spätere Nutzung archivieren zu können.

Hier zeigte sich eine Parallele zur Arbeit von Verwaltungsarchiven, die ebenfalls darauf angewiesen sind, dass der Entstehungskontext von Unterlagen dokumentiert wird. Es ist daher kein Zufall, dass im Umfeld der öffentlichen Archive ein Werkzeug gefunden wurde, das einen großen Teil der gewünschten Funktionen bereits besaß. Die Schweizer Firma Docuteam bot mit den Komponenten docuteam packer und docuteam feeder bereits die wesentlichen Bausteine für diese Aufgabe als Open Source Software an.<sup>13</sup> Forschende und abliefernde Stellen arbeiten vor Ort direkt mit dem docuteam packer, während der docuteam feeder zentral im Hintergrund betrieben wird.<sup>14</sup>

#### **Abb. 7:** Arbeitsablauf mit docuteam packer

Aus Sicht der Langzeitarchivierung laufen die folgenden Schritte ab (Abb. 7): Zunächst werden Daten lokal in den eigenen Arbeitsumgebungen der Forschenden in den docuteam packer importiert, in einer Verzeichnisstruktur abgelegt und Metadaten mit der gewünschten und vereinbarten Granularität erfasst. Dieser Prozess kann z.B. über die ganze Dauer eines Forschungsprojekts fortgesetzt werden. Nach Abschluss der Erfassung wird die Übergabe, die sogenannte Submission ausgelöst, die die Erzeugung eines Submission Information Packages (SIP) und seine Übergabe an den docuteam feeder bewirkt. Hier liegt ein zentraler Vorteil der Lösung: Kunden müssen sich nicht selbst mit den Anforderungen an SIPs auseinandersetzen und können trotzdem aufgrund der hinterlegten Konfiguration den internationalen Standards entsprechende SIPs erstellen.

Der docuteam feeder verrichtet seine Arbeit im Hintergrund. Stark vereinfacht kann man sagen, dass der feeder das SIP so aufbereitet, dass es dem ETH Data Archive übergeben werden kann. In der Praxis übernimmt der docuteam feeder eine zentrale Rolle während der Ingest-Phase. Nachdem gewisse Erweiterungen umgesetzt wurden, wird er an der ETH-Bibliothek auch intensiv für den Ingest weiterer Daten ins ETH Data Archive genutzt, um z.B. die stapelweise Verarbeitung von Digitalisaten der ETH-Bibliothek zu steuern. Diese Prozesse werden nicht vom docuteam packer angestoßen.

---

<sup>13</sup> Zunächst noch unter den Namen docupack und documill.

<sup>14</sup> <https://wiki.docuteam.ch/doku.php?id=docuteam:oais>.

Für die ETH-Bibliothek war von Anfang an die Aussicht vielversprechend, mit diesen Tools sowohl Forschende als auch das Hochschularchiv der ETH Zürich bedienen zu können. Für beide Anwendungsfälle wurden umfangreiche Anpassungen umgesetzt. Zwei Nachteile bzw. Limitierungen blieben bestehen: Für größere Datenmengen ist die Verarbeitung im docuteam packer durch die im Hintergrund ablaufenden Prozesse wie Prüfsummenberechnung und verschiedene Kopiervorgänge zu langsam. Obwohl es keine harte technische Grenze gibt, sollten Nutzer, aus Gründen der praktischen Handhabung nicht mehr als zwei Gigabyte Daten pro Paket laden. Diese Einschränkung ist für den Einsatz in vielen Forschungsgruppen inakzeptabel. Einige Forschungsgruppen haben sich dennoch damit arrangiert, da der Nutzen für bestimmte Anwendungsfälle überwiegt.

Die zweite wesentliche Limitierung liegt in der Skalierung aus betrieblicher Sicht. Der docuteam packer muss zwar nicht installiert werden, aber vor Ort in der Arbeitsumgebung der Kunden zum Laufen gebracht werden und zwar in der Regel so, dass nicht alle Mitarbeitenden Zugriff auf die Daten haben. Dieser Aufwand ist für einen überschaubaren Kundenkreis zwar machbar, wäre aber für eine Vielzahl von Gruppen nicht zu leisten.

Diese Einschränkungen haben dazu geführt, dass der docuteam packer trotz seiner Vorteile nicht generell für den Einsatz bei Forschungsgruppen empfohlen wird, sondern eher bei Bedarf ins Gespräch gebracht wird, wenn Alternativen für den konkreten Anwendungsfall nicht zielführend zu sein scheinen. In der Regel genügen die Workflows, die in der Research Collection eingebaut sind, den meisten Kundenbedürfnissen völlig.

Im Hochschularchiv der ETH Zürich ist der Einsatz des docuteam packer weiterhin unbestritten. Ob der docuteam packer für weitere Forschungsgruppen in Frage kommt, wird davon abhängen, ob sich die Performance verbessern lässt. Überlegungen dazu sind im Gange.

#### **4 Situation in der Schweiz**

Da ein effektives und effizientes Datenmanagement sowohl für Forschende als auch für Informationsspezialisten über alle Institutionen hinweg immer anspruchsvoller wird, stellte sich die Frage, wie man sich auf nationaler Ebene am besten organisieren und gegenseitig unterstützen kann.

Das Projekt „Research Data LifeCycle Management: From Pilot Implementations to National Services (DLCM)“ zielte darauf ab, die Bemühungen von acht Schweizer Hochschulen (EPFL,<sup>15</sup> ETH Zürich,<sup>16</sup> Fachhochschule Westschweiz,<sup>17</sup> Universität Basel,<sup>18</sup> Universität Genf,<sup>19</sup> Universität Lausanne,<sup>20</sup>

---

<sup>15</sup> <https://www.epfl.ch>.

<sup>16</sup> <https://www.ethz.ch>.

<sup>17</sup> <https://www.hes-so.ch>.



Universität Zürich<sup>21</sup>), vertreten durch ihre Bibliotheken und IT-Abteilungen zu vereinen, einschließlich des bereits bestehenden nationalen Dienstleisters für Hochschulen SWITCH.<sup>22</sup> Das Projekt wurde im Auftrag der ehemaligen Schweizerischen Universitätskonferenz (SUK, heute swissuniversities<sup>23</sup>) im Rahmen des Programms SUK P-2<sup>24</sup> initialisiert, das zum Ziel hatte, den Umgang mit wissenschaftlichen Informationen in der ganzen Schweiz zu verbessern.

Mit dem Start von DLCM im Jahr 2015 haben v.a. die Bibliotheken der ETH Zürich und der EPFL ihre bereits seit 2012 bestehenden Bemühungen im Bereich Forschungsdatenmanagement intensiviert und gemeinsam an der Erarbeitung von Dienstleistungen im Bereich Datenmanagementpläne (DMP) und Forschungsdatenmanagement gearbeitet.

Die Dienstleistungen und Trainings wurden in enger Zusammenarbeit mit den wissenschaftlichen IT-Dienstleistern und dem Stab Forschung der jeweiligen Hochschulen entwickelt. Dies ermöglichte es den jeweiligen Hochschulbibliotheken, auf die Bedürfnisse der Forschenden zugeschnittene Angebote anzubieten und den gesamten Datenlebenszyklus abzudecken.

Dabei waren die Bibliotheken der ETH Zürich und der EPFL federführend an der Erarbeitung einer Datenmanagement-Checkliste, einer Wegleitung zum Ausfüllen der DMPs für den Schweizerischen Nationalfonds (SNF), sowie der Etablierung von Trainings und Beratungsdienstleistungen (siehe auch Kapitel Kurse und Beratungen) beteiligt.

Die Erfahrungen aus dem DLCM Projekt haben bestätigt, dass es unbedingt lokale Partner braucht, bei denen die Forschenden an ihrer eigenen Hochschule anfragen oder vorbeigehen können, um eine persönliche Beratung zu erhalten. Die Hemmschwelle für eine Beratung durch Fachstellen anderer Universitäten liegt relativ hoch. Zum einen birgt die physische Distanz hohe Zeit- und allenfalls finanzielle Kosten (z.B. Reisekosten oder Teilnahmegebühren für externe Kursteilnehmer).

Insbesondere Doktorierende haben oft nicht die Möglichkeit, solche Beratungen als Arbeitszeit zu verbuchen oder sich die Reise- und Teilnahmegebühren rückerstatten zu lassen. Zum anderen verfügen die lokalen Partner über die beste Kenntnis der Gegebenheiten vor Ort und können zielgerichtete Hinweise auf Möglichkeiten für weiteren Support an der Hochschule des jeweiligen Forschenden geben.

---

<sup>18</sup> <https://www.unibas.ch>.

<sup>19</sup> <http://www.unige.ch>.

<sup>20</sup> <http://www.unil.ch>.

<sup>21</sup> <https://www.uzh.ch>.

<sup>22</sup> <https://www.switch.ch>.

<sup>23</sup> <https://www.swissuniversities.ch>.

<sup>24</sup> Inzwischen umbenannt zu „swissuniversities P-5“.

## 5 Ausblick

Die ETH Zürich verfügt zwar seit Jahren über eigene Richtlinien für Integrität in der Forschung, doch fehlt bisher eine eigentliche Datenpolicy. Die Hochschule hat bisher den Weg verfolgt, durch den Aufbau bedarfsgerechter Angebote Lösungen für die Herausforderungen des Datenmanagements zu bieten. Angesichts der öffentlichen und Fachdiskussionen sowie der Anforderungen seitens der Geldgeber ist besonders bei jüngeren Forschenden an der ETH Zürich der Bedarf gewachsen, die Erwartungen der Hochschule an ihre Angehörigen zum Datenmanagement und zur Bereitstellung von Daten für Dritte zu klären. Interne Diskussionen zu einer Datenpolicy und / oder zu Empfehlungen zum Umgang mit Forschungsdaten laufen bereits seit einiger Zeit.

Datenmanagement ist Teil des Forschungsprozesses und seiner Qualitätssicherung. Eine unnötige Bürokratisierung des Forschungsprozesses gilt es jedoch zu vermeiden. Stattdessen sollten Forschende möglichst schon in ihrer Fachausbildung die nötigen Kompetenzen erwerben, um ein angemessenes Datenmanagement selbst umzusetzen. Dies ist auf die Dauer schlanker, günstiger und nachhaltiger.

Der gegenwärtige Aufwand für Beratungen und Schulungen mag überschaubar erscheinen, erfordert aber bereits erhebliche Ressourcen. Sollten tatsächlich alle 530 Professoren (ETH Zürich 2017) der ETH Zürich in absehbarer Zeit eine Beratung und Weiterbildung wünschen, müssten die Personalressourcen stark erhöht werden.

Die Verankerung des Forschungsdatenmanagements in der Ausbildung jedes Studiengangs wäre ein großer Schritt in die richtige Richtung. Es wird jedoch nur möglich sein, das Kursangebot langfristig zu reduzieren, wenn international eine vergleichbare Ausbildung zum Datenmanagement während des Fachstudiums angeboten wird, denn insbesondere beim Übergang zu Masterstudium und Doktorat ist die Mobilität zwischen den Hochschulen hoch. Neben der Bereitstellung der technischen Infrastruktur wird aber weiterhin die individuelle Begleitung seitens der jeweiligen Hochschulbibliothek ein wichtiger Teil der Forschungsunterstützung bleiben.

Die ETH-Bibliothek wird sich verstärkt bemühen, das Thema Forschungsdatenmanagement möglichst früh in der wissenschaftlichen Karriere zu verankern und auch Studierende für die Thematik zu sensibilisieren. Dazu gehört, dass die ETH-Bibliothek für den Sommer 2019 eine Summer School vorbereitet. Während vier Tagen sollen junge Forschende aus dem ETH-Bereich die Prinzipien des Forschungsdatenmanagements entlang des Datenlebenszyklus kennenlernen und in praktischen Beispielen und Übungen selbst vertiefen. Eine wesentliche Aufgabe dabei wird sein, die eigene Arbeitsweise und die eigenen Daten zu hinterfragen und beispielsweise einen sinnvollen Datenmanagementplan zu erstellen. Der Besuch der Summer School wird mit ECTS-Kreditpunkten

honoriert, was besonders für Doktorierende attraktiv ist, da sie gemäß ETH-Richtlinien während des Doktoratsstudiums auch Kreditpunkte außerhalb ihrer spezifischen Fachdisziplin sammeln müssen.

Die ETH-Bibliothek strebt zudem an, Beratungsangebote zu entwickeln, die speziell auf die Bedürfnisse von Projektleitenden zugeschnitten sind. Aus langjähriger Erfahrung und Austausch mit Kollegen wissen wir, dass Datenmanagement nur funktionieren kann, wenn es von der Leitungsebene getragen wird. Auch gemäss den Richtlinien für Integrität in der Forschung der ETH Zürich (ETH Zürich 2011) liegt es in der Verantwortung der Leitungspersonen von Forschungsprojekten dafür zu sorgen, dass das Datenmanagement innerhalb der Teams funktioniert und alle Teammitglieder informiert sind.

Bei all diesen Angeboten zur Forschungsunterstützung ist von Seiten der Hochschulbibliotheken und IT-Services neben fachlichen Kompetenzen nicht zuletzt auch Fingerspitzengefühl gefragt. Keinesfalls sollte der Eindruck entstehen, dass die Forschenden bevormundet werden. Ziel muss es sein, den Forschenden den Zugang zu den Methoden und Tools zu erleichtern, die ihnen bei der Wahrnehmung ihrer Aufgaben helfen können. Das Hauptinteresse der Forschenden liegt darin, exzellente Forschung mit einem adäquaten Datenmanagement zu untermauern – und nicht umgekehrt. Es ist also zentral, zwischen den Vorgaben zum Datenmanagement und den konkreten Bedürfnissen der Forschenden zu vermitteln. Dieser Balanceakt wird alle, die Forschende beim Management ihrer Daten unterstützen, in den kommenden Jahren herausfordern.

## Literaturverzeichnis

- Barillari, Caterina, Diana S M Ottoz, Juan Mariano Fuentes-Serna, Chandrasekhar Ramakrishnan, Bernd Rinn, and Fabian Rudolf. 2016. "OpenBIS ELN-LIMS: An Open-Source Database for Academic Laboratories." *Bioinformatics* 32 (4): 638–40. doi:10.1093/bioinformatics/btv606.
- Bornmann, Lutz, and Rüdiger Mutz. 2015. "Growth Rates of Modern Science : A Bibliometric Analysis Based on the Number of Publications And." *Journal of the Association for Information Science and Technology* 66: 2215–22. doi:10.1002/asi.23329.
- ETH Zürich. 2011. "Richtlinien Für Integrität in Der Forschung - Guidelines for Research Integrity." ETH Zürich. doi:10.3929/ethz-b-000179298.
- . 2015. *Reglement Für Das Archiv Der ETH Zürich*. <https://rechtssammlung.sp.ethz.ch/Dokumente/420.1.pdf>.
- . 2017. "Personenkennzahlen." Zürich. [https://www.ethz.ch/content/dam/ethz/main/eth-zurich/ArbeitenLehrenundForschen/Personalkennzahlen/Files/Personalkennzahlen\\_2017.pdf](https://www.ethz.ch/content/dam/ethz/main/eth-zurich/ArbeitenLehrenundForschen/Personalkennzahlen/Files/Personalkennzahlen_2017.pdf).

———. 2018. “ETH Scientific IT Services.” <https://sis.id.ethz.ch>.

European Commission. 2018. “Open Access & Data Management.” *Participant Portal H2020 Online Manual*. [http://ec.europa.eu/research/participants/docs/h2020-funding-guide/cross-cutting-issues/open-access-dissemination\\_en.htm](http://ec.europa.eu/research/participants/docs/h2020-funding-guide/cross-cutting-issues/open-access-dissemination_en.htm).

Hirschmann, Barbara. 2018. “Die Research Collection der ETH Zürich.” *ABI Technik* 38 (3): 223–33. doi:10.1515/abitech-2018-3003.

Noorden, Richard Van. 2014. “Global Scientific Output Doubles Every Nine Years.” *Nature News Blog*. <http://blogs.nature.com/news/2014/05/global-scientific-output-doubles-every-nine-years.html>.

Schweizerische Eidgenossenschaft. 1999. *Bundesgesetz Über Die Archivierung*. Schweizerische Eidgenossenschaft. <https://www.admin.ch/opc/de/official-compilation/1999/2243.pdf>.

Schweizerischer Nationalfonds. 2017. “Open Research Data.” [http://www.snf.ch/de/derSnf/forschungspolitische\\_positionen/open\\_research\\_data/Seiten/default.aspx](http://www.snf.ch/de/derSnf/forschungspolitische_positionen/open_research_data/Seiten/default.aspx).

Sesartic, Ana. 2017. “Surfen im Datenmeer. Datenspeicherung und Schutz vor Datenverfall. Datenspeicherung und Schutz vor Datenverfall.” *Explora*. ETH Zürich. doi:10.22010/ethz-explora-0002-de.

Sesartić, Ana, Andreas Fischlin, and Matthias Töwe. 2016. “Towards Narrowing the Curation Gap—Theoretical Considerations and Lessons Learned from Decades of Practice.” *ISPRS International Journal of Geo-Information* 5 (6): 91. doi:10.3390/ijgi5060091.

Springer Nature. 2018. “Research Data Policies FAQs.” <https://www.springernature.com/gp/authors/research-data-policy/faqs/12327154>.

Wallis, Jillian C., Elizabeth Rolando, and Christine L. Borgman. 2013. “If We Share Data, Will Anyone Use Them? Data Sharing and Reuse in the Long Tail of Science and Technology.” *PLoS ONE* 8 (7). doi:10.1371/journal.pone.0067332.

Ana Sesartic Petrus  
ETH-Bibliothek, ETH Zürich

CH-8092 Zürich

Schweiz

[ana.petrus@library.ethz.ch](mailto:ana.petrus@library.ethz.ch)

Matthias Töwe

ETH-Bibliothek, ETH Zürich

CH-8092 Zürich

Schweiz

[matthias.toewe@library.ethz.ch](mailto:matthias.toewe@library.ethz.ch)